Office for National Statistics

Teaching Excellence Framework: Review of Data Sources – Interim Report

Methodology Advisory Service
Office for National Statistics
February 2016

Office for National Statistics

Contents

1 Introduction

1.1 Overview and Context

The Government has proposed the introduction of a Teaching Excellence Framework (TEF) with the underlying aim of identifying, rewarding and encouraging the highest quality of teaching within Higher Education Institutions (HEIs). It also aims to benefit students, offering a wider range of courses to suit their needs and to better ensure graduate employability while ensuring that diversity and social mobility are accounted for across all institutions. It is proposed that, from the start of the 2017/18 academic year, assessments will be made by an expert panel drawing on a set of metrics related to teaching quality and student outcomes together with information supplied by the Educational Institutions [1].

Three data sources have been identified to underpin the metrics:

- student record data collected by the Higher Education Statistics Agency (HESA) from individual institutions and individual learner data for FE colleges
- HESA UK performance indicators, based on returns from the Destination of Leavers from Higher Education Survey (DLHE) which is often collated by careers services departments within institutions and
- student satisfaction indicators derived from the National Survey of Students (NSS) which is commissioned by the Higher Education Funding Council for England (HEFCE) and administered by Ipsos-MORI

In order to help the Government design a robust assessment within the TEF, the Office for National Statistics (ONS) was asked in November 2015 to carry out a review of these data sources. The Higher Education Analysis Division within BIS is acting as sponsors for this work on behalf of the Minister of State for Universities and Science.

1.2 Specification of the Review

The specification for the review was set out by the customer at the start of the work; it is summarised here.

Objectives

ONS will provide and independent and expert assessment of

- the quality and robustness of the sources of information the Government is proposing to use in the TEF assessment process from 2017/18 onwards
- what implications this might have in relation to making clear and robust determinations on an institution's or course's performance against the purpose for which that metric has been chosen
- where areas of improvement are identified, the extent to which these might be addressed either within the existing survey or via alternative means

Scope

The review is to provide as assessment of the sources of data used to calculate the proposed metrics (as set out in the Higher Education Green Paper). The sources are:

- DLHE collected by HESA for HEIs and HEFCE for FECs
- continuation of students between first and second year of study (from the UK performance indicators published by HESA)

- student satisfaction indicators on “Teaching on my Course”, “Assessment and Feedback” and “Academic Support” (based on questions from the NSS, commission by HEFCE)

The topics expected to be covered include: collection methods, representivity, response rates, weighting, the potential for bias and scope to make inferences for sub-groups of the student population.

External Engagement

ONS is expected to contact and interview organisations and individuals responsible for the underlying data sources and potentially the research organisations involved in the collection, design and collation of the data.

Deliverable and Timing

The deliverable is an assessment report with recommendations.

1.3 Approach

A small review team has been assembled from within the methodology group at ONS. It comprises two data collection methodologists, one methodologist with wide statistical process experience and the head of the Methodology Advisory Service providing management of the work.

In carrying out this review, the review team has assessed the data sources against standard quality frameworks and general statistical good practice pertaining to Official Statistics. ONS publishes guidelines for measuring statistical quality [2]. The generic statistical business process model is also a helpful guide and is used in the ONS Regular Quality Review process which is one method ONS uses to asses and improve statistical outputs [3], [4]. Assessing the data sources against quality measures for Official Statistics can be considered as “setting a high bar”.

The review undertaken is not exhaustive. A full review of the data sources would require a much longer period, so the review team has been selective in examining the key areas to determine robustness within a timescale that supports the Government’s development timetable for TEF. In line with the specification, the review team has interviewed all the parties involved in the underlying data sources; this includes staff from the following:

- HEFCE - Higher Education Funding Council for England
- HESA – Higher Education Statistics Agency
- HESPA – Higher Education Strategic Planners Association
- Ipsos-MORI

In addition, the review team carried out interviews with staff from two higher education institutions; it is acknowledged that this cannot be considered a representative sample of the sector as a whole.

1.4 Acknowledgements

The ONS review team would like to thank staff from all of the organisations listed above for their input into this brief study, but particularly from HEFCE and HESA who have been extremely helpful.

1.5 Structure of this report

Chapter 1 starts by summarising the context, scope and objectives of the review, the approach the review team has taken and the quality standards against which an assessment has been made. Data collection is

considered in chapter 2 and statistical processing is assessed in chapter 3. General conclusions are described in chapter 4 and recommendations presented. Appendix A describes the methodology and processes behind the cognitive interviewing process; this provides background explanatory material for chapter 2

1.6 Interim version of the report

Not all review work is complete; further work is needed in two areas: questionnaire design and the use of benchmarks.

2 Review of data collection

This section of the report looks at each of the data collection instruments from a qualitative perspective. In the first instance expert reviews have been carried out on both the DLHE and the NSS. In addition, a review of the collection process undertaken as part of the HESA data collection procedure for the student record has also taken place with cognitive interviews being undertaken with staff within institutions responsible for the delivery of data to HESA. Respondents in all three areas were asked about the ease with which the data collection process can be completed and were asked to voice any concerns they had with the process itself or with the data collected.

Sample selection and composition

With a very limited period to complete the review, purposive sampling techniques were employed. The sample is designed to pick a small number of cases that will give the most information on the topic of interest [5]. Respondents are identified from appropriate sub-populations taking into account characteristics of interest [6]. The goal is not to generalise to the population, but to gain insight into 'phenomenon, individuals or events' [7].

For expediency an opportunity sample was used. Once each of the areas to be investigated had been agreed, a list of contacts was compiled. These came either through suggestions from HEFCE or via contacts within ONS. There were three main purposes for testing:

- to gauge the relative ease of completion for each collection instrument
- to gauge whether there are concerns with regards to the quality of the datasets compiled
- to establish levels of confidence in the use of the above data sources to create performance indicators

Respondents were selected from two higher education institutions and Ipsos-MORI. The methodology and processes behind the cognitive interviewing process is described in Appendix A.

2.1 Review of the National Student Survey

2.1.1 Methodology

The NSS is a key source of data about student experiences. It was designed to inform student choice and contribute to public accountability; it is also used by the media as a data source for university league tables. It gathers information from students on their experiences of their courses before they graduate, in most cases in their final year. The contact list is produced from the HESA student record; this is intended to be a complete list of suitable students, so the study is aiming to be a census. Ipsos-MORI is contracted to administer the data collection. Recruitment to take part in the survey is primarily linked to the term time address of the student; however, if no contact is made, notification will be sent to the out of term address. The list of students can be augmented with academic institutions having the ability to either add or remove respondents from the student list as necessary.

A review of the collection instrument will be carried out and the results will appear in the final version of the report, in section 2.1.2. An interview with Ipsos-MORI, the organisation responsible for collecting the data, is summarised in section 2.1.3. The previous NatCen review is referenced in 2.1.4 and overall findings are set out in section 2.1.5.

2.1.2 Questionnaire design

This aspect of the review work has not been completed; it will be reported in the final version of the report. The work will be informed by the appropriate literature [8-10].

2.1.3 Summary of interview with Ipsos-MORI

The review team spoke to a member of staff at Ipsos-MORI who is familiar with the NSS.
Coverage
There is over-coverage of some groups; for example, students on three year courses. There is also under-coverage however; for example, minority ethnic groups and no coverage of students who leave prior to their final year.

It was noted that HEFCE are proposing changes to the NSS post-2017. HEFCE have reported that the proposed longer-term changes to the NSS will address the need to include data from a range of students. In particular, mature students and those from minority ethnic groups; it recognised that they are disproportionately represented within the non-responder group. HEFCE also proposes to examine expanding the student base by including those who withdraw from their course and those who are studying at alternative institutions [10]. It was felt that these changes should be implemented as soon as is possible to improve the quality of information being considered by the panel.

There is also the issue of those students who do not continue into the final year. Although there are many possible reasons behind such events, it is not unreasonable to conclude that some of this group might have been dissatisfied with their student experience, and so their input could be a key piece of information for the panel.

Response

Although NSS response rates can be considered to be good at around 70%, the non-responders are still of interest. The representative of Ipsos-MORI explained that the survey's non-responders are very similar to the late responders; this could help to account for non-responders.

2.1.4 NatCen Questionnaire Review

In 2014, NatCen presented a review of the NSS at the behest of the UK Higher Education Funding Bodies. This review will not repeat their findings regarding the questionnaire as there is overall agreement and their conclusions are well-documented [11].

2.1.5 Overall Summary of findings

The NSS is a key source of data about student experiences. It was designed to inform student choice and contribute to public accountability; it is also used by the media as a data source for university league tables. In examining the survey, there are certain issues around collection of the data which HEFCE should be aware of when considering the use of NSS data as part of quality indicators:

- under-reporting of certain groups and over-coverage of others is a matter of concern and could lead to bias in use of the data
- the lack of a voice from those who did not complete their course is a potential weakness in the planned quality indicators; neither DLHE nor NSS have this element. As for the previous point, this could result in bias. While the NSS response rates are good by modern standards, an understanding of the non-responders would be of significant benefit

2.2 Review of the Destination of Leavers from Higher Education

2.2.1 Methodology

Coverage

Students in the DLHE sample are those who have completed their course. Research around user-satisfaction and the destination of students would need to balance the views of those who have completed their courses with those who have not. Those who have not continued will have done so for many different reasons but within this group it is highly likely that there will be students who were not satisfied with their student experience. While this survey looks at outcomes for successful leavers; it should be noted that the continuation of students is included as a metric.

Non-response

Response rates to the DLHE are very good, despite the current climate of falling response rates generally. However, this does not mean that non-response can be neglected. Even with relatively good response rates, significant bias can be present.

Questionnaire administration

This aspect of the review work has not been completed; it will be reported in the final version of the report.

2.2.2 Questionnaire Design

This aspect of the review work has not been completed; it will be reported in the final version of the report.

2.2.3 Summary of findings

- the survey is concerned with the destination of students following completion of their course; outcomes for students who were not successful are not captured in this source (nor in the NSS). An expression of dissatisfaction is captured, however, in the metric concerned with non-continuation rates
- an understanding of non-responders is important; without this there is scope for bias

2.3 Review of the HESA data collection process for the student record

HESA is responsible for gathering student data from universities, higher education colleges and other providers of higher education. Once amalgamated, the data is used to inform funding models, which are implemented by Government and higher education funding bodies. The same data form the basis for statistics which can be accessed by a range of potential users including prospective students, academic researchers, professional bodies and the media; bespoke information can also be provided on request.

The data collected is comprehensive, comprising about 200 pieces of information for each student. This is combined into the “student record”. Administration teams within academic institutions are responsible for ensuring that the required data is transmitted back to HESA within a given series of deadlines, which incorporate additional time for the cleaning of data and amendments where necessary. The data is returned in XML format according to HESA criteria, which is transmitted through a secure portal to HESA directly. As part of the evaluation process ONS contacted several HE providers to gain further insight into the collection process from their perspective and to gauge their opinion on the quality of the data that resulted from the data collection process.

Summary of findings

This section presents a brief summary of responses from two HEIs.

The staff involved in collecting the data for the HESA return reported that it is a substantial activity that takes a significant amount of time to ensure the data meets the specific requirements. While some of the data is collected to meet the needs of the institutions, a significant amount is only required for the HESA return. The quality of the return is helped by restrictions on the values of data that can be returned and by cross validation checks.

Some technical issues were described relating to specific situations of students and how they should be reported. There was a concern that there was potential for inconsistency between the ways institutions reported student circumstances. The collection process does include mechanisms to reduce the chance of this. To ensure a high degree of consistency between returns from institutions, HESA publishes coding guidance and provides a support service to institutions to resolve queries.

Detailed cognitive testing of the collection process would be needed to investigate whether consistency issues were significant and this could not be included in the review.

2.3.1 Previous ONS Analysis of HESA Data

The HESA student record data were examined by the "Beyond 2011 Programme" in 2013 as a potential source of administrative data to support an alternative approach to the 2021 Census [12]. The main conclusion was that the extensive checking carried out by HESA ensured that the data was of high quality. A comparison was made between the 2010/11 HESA data and the 2011 Census showed a good degree of agreement between the age, sex and geographical distribution of students across England and Wales.

2.3.2 Additional feedback

During the consultations, respondents expressed reservations about wider issues related to the use of information from the NSS and the DLHE. Concerns included:

- limited variation between institutions of the raw scores from the student responses
- difficulty in trying to compare widely differing institutions
- difficulty in capturing the wider benefits beyond academic results of attending a higher education institution

2.3.3 Closing remarks

Consulting only two HEIs doesn't provide a representative sample of the overall population, of course; this was not the intention of this part of the review work. They need to be included as they are important constituents of the overall process.

3 Statistical Processing

3.1 Coverage

The coverage of the NSS and DLHE surveys are clearly defined in terms of which students should be contacted. Similarly, the data for students returned by Higher Education Institutions for the student record is also clearly specified.

The DLHE is targeted at all students reported to HESA for the period 01 August to 31 July as obtaining relevant higher education qualifications and whose study was full-time or part-time. The target population includes students from all domiciles, including Guernsey, Jersey and the Isle of Man. Excluded from the target population are leavers with further education level qualifications, leavers who studied mainly overseas, incoming exchange students, students who are on an intercalated course, and deceased students.

The NSS population is determined by a HESA target list, which identifies undergraduate students expected to be in the final year of their course in the period covered by the survey.

While the coverage of these sources is clear, this is not necessarily the coverage required for the purposes of the TEF. The review team is not clear what the target coverage is for the TEF and this is an important question that must be decided in order to assess how well the HESA record, the DLHE and the NSS meet the target. This question has been discussed with the customer during the review, but has not been answered during the period of this review. Without this information, the review team cannot assess the degree of coverage provided by the student record, the DLHE and the NSS.

It is possible that the surveys do not cover all students that are required (undercoverage) and might include some students that are not in scope of TEF (overcoverage). Any coverage error of this nature has the potential to introduce bias into the TEF outcomes due to the surveys not properly representing the target population. Without knowing the target population, it is impossible to comment on the likely impact of such a bias. Once the target population has been defined, it is recommended that consideration is given to any important undercoverage of the surveys. Any overcoverage should be dealt with by removing data for those students that are not in scope of TEF. Clearly not removing overcoverage can also introduce bias into results.

In other situations, it has been helpful to consider two target populations. Firstly, a conceptual, or theoretical, target population is defined; this is the ideal situation and would contain all the students and student circumstances that would fully support the metrics. In practice, this may be unachievable. Next define a practical target population – this is not ideal but allows for reasonable compromises over what is desired and what can be achieved. Of course, it should be as close to the theoretical target as possible. Then compare the current coverage of the HESA student record, the DLHE and the NSS to the practical target and identify any gaps. If the gaps can be closed then effort should be made to do so; if they can’t, a bias is possible and ways to estimate this bias should be sought.

3.2 Response

For the NSS, Ipsos-MORI uses a focused response chasing strategy. Students who don't respond within a given period, normally three weeks, are followed up using telephone and post. This process is carried out until minimum response rates are met for each publication breakdown. The response target in each case is 50% and at least 10 responses. HESA provides guidance for HEIs regarding the collection of the DLHE – four methods are offered, telephone, paper, web and electronic form by email. At the overall level the most

recent response rates stated for the surveys are 71% for NSS and 79% for DLHE. The response rate for the DLHE is lower for non-EU domiciled leavers.

It is assumed that the response data are fully representative of the whole population. For both surveys, there is no non-response weighting or imputation carried out to adjust for non-response. In general, it is common for non-respondents to surveys to have different characteristics to people that respond. If there are differences with respect to the TEF outcomes, not treating non-response could lead to bias in the metrics.

A previous study on non-respondents to the NSS followed a cohort of first degree students to see whether they were invited to take the NSS [10]. The analysis showed that young students were more likely to be invited than mature students and white students were more likely to be invited than students from other ethnic backgrounds. In 69% of these cases students were correctly excluded due to not meeting the criteria for the survey. There were also a significant proportion of students invited to complete the survey either in the wrong year or when they did not qualify at all, suggesting the potential for overcoverage.

The level of non-response to both surveys is easily high enough to suggest the possibility of non-response bias, especially given the lack of treatment to deal with non-response. The study on non-respondents to the NSS suggests that this may be exacerbated by a lower chance of being invited to take part in the survey for particular sub-groups of the population.

The HESA student record data provide an opportunity to examine the characteristics of responders and non-responders. If there is little difference, this could be considered as evidence that non-response is not introducing bias. Alternatively, differences in characteristics would be evidence that bias may be being introduced and specific analysis would be required.
It was possible to gain some insight into the characteristics of non-responders for this review thanks to a visit to HEFCE. With assistance from HEFCE staff, response rates were inspected for a variety of key population breakdowns in the NSS data. Details of this analysis are described below.

3.2.1 Non-response analysis

The main analysis of non-response was based on inspecting 2015 NSS data, broken down by various sub-groups. The robustness of this analysis was checked by comparing the same analyses on 2009 data and also by looking at a time series of NSS data from 2005 to 2015. The overall response levels were lower in previous years, but the response patterns were mostly similar.

Response was examined at the overall (unit) level and at the item level for Question 22, which is the key question on the survey (relating to students' overall satisfaction with the quality of the course).

Broad ethnicity

In 2015, unit response rates for the broad ethnicity grouping are at a similar level for all categories. However, White students (72%) were slightly more likely to respond than Asian (70%), Mixed ethnicity (70%) and Black (67%) students. A similar pattern was observed in other years.

Age

The Under 21 age group had a response rate around 7% higher than that of other age groups in 2015 and generally had the highest response in the other years examined.
Socio-economic classification

There is lower response for the most deprived Socio-economic class: category 8 (never worked / long term unemployed). Category 8 has around 69% unit response, whereas categories 1 to 7 are all between 73% and 75% response. This variable was not available for the earlier years examined.

The Socio-economic classification in NSS is self-classified. A more robust, related measure is POLAR (Participation of Local Areas). The scale for POLAR ranges from category 5 for areas with a high percentage of people going into higher education down to category 1 for areas with a low percentage going into higher education. This variable also shows slightly lower response for more deprived areas, 71% for category 1 areas rising up to 74% response for category 5 areas. The same pattern was observed in other years.

Domicile

Response rates are a little lower for students coming from outside of the UK. Response for the category Rest of the world was 67%, compared with 70% for EU countries outside of the UK. Response for the constituent UK nations varied from 71% to 74%. This pattern was also observed in other years.

Part-time

Response was considerably lower for part-time students (60%) compared with full-time (72%). Only 4% of overall response comes from part-time students. The number of part-time students has halved since 2009, but the discrepancy in response rates between part-time and full-time was consistent over the period investigated.

Disability

In 2015, the response rates were fairly uniform comparing the Declared disability with Dyslexia and No known disability categories. However, looking over the period 2005 to 2015, overall response for the Dyslexia and No known disability categories has increased, whereas for the Declared disability category it has stayed relatively level.

Sex

Females have higher response rates to NSS than Males. In 2015, 75% of Females responded compared to 69% of Males. The same pattern is observed in all the years investigated.

Institution (UKPRN)

There is considerable variation in response between institutions. Most vary between 55% and 80%. Some of the institutions with lower responses have a very small number of students. But note that there are some larger institutions with very low response rates. For example, response from Oxford University in 2015 was only 55%.

The key result is that a degree of differential response was observed for a number of population characteristics, including Ethnicity, Sex, Age and Socio-economic Classification. The difference in response rate between some of the categories for these characteristics was generally around 5%. While this is a relatively small difference, it is certainly large enough to introduce bias into results. The impact of any bias could be greatly reduced through the introduction of appropriate imputation or non-response weighting techniques.

3.2.2 Data Quality

Respondent error is a common problem in all types of survey. Some types of respondent error are random and can be assumed to approximately cancel each other out. Respondent bias occurs when those errors are systematic. In this case, the errors can lead to unrepresentative statistics if left untreated in the data. Edit rules are generally used to detect and correct important errors in survey responses.

The NSS uses the following edit rules:

- Invalid language or individuals are removed.
- Invalid responses are flagged and treated as non-response. The following cases are defined as invalid:
- Multiple responses to the same question (if there are more than four questions with multiple responses, the whole response is marked as invalid)
- Not applicable ("N/A") responses
- If a whole section of questions either contains multiple responses or "N/A" responses, the whole response is marked as invalid

Given that the survey consists only of attitudinal tick box questions, this should be sufficient to ensure reasonable data quality.

3.2.3 The Potential for Inference

In order to use data from DLHE and NSS to make valid inferences about higher education institutions, it is necessary for those data to be of sufficient quality and to be able to compare meaningfully between institutions, or to compare institutions to a derived sector statistic.

It is instructive to carry out a basic examination of the data and its variation; the review examined published NSS data to understand if it is possible to distinguish between institutions at the overall level. Confidence intervals are not available for more detailed breakdowns of the NSS data, but the sample sizes of smaller domain estimates have been examined to give a general indication of precision. Equivalent data for DLHE was not available for this review, but ideally the same type of analyses should be carried out.

For published NSS proportions, confidence intervals are calculated using the Wilson score method. Figure 1 below shows, for each institution, the percentage of students agreeing with the statement that “overall they are satisfied with their course”, question 22 on the NSS. The institutions are shown in ascending order of the outcome variable. Each estimate is accompanied by its confidence interval, shown as bars on the graph.

Figure 1: Percentage of students with positive response to Q22 – Overall satisfaction

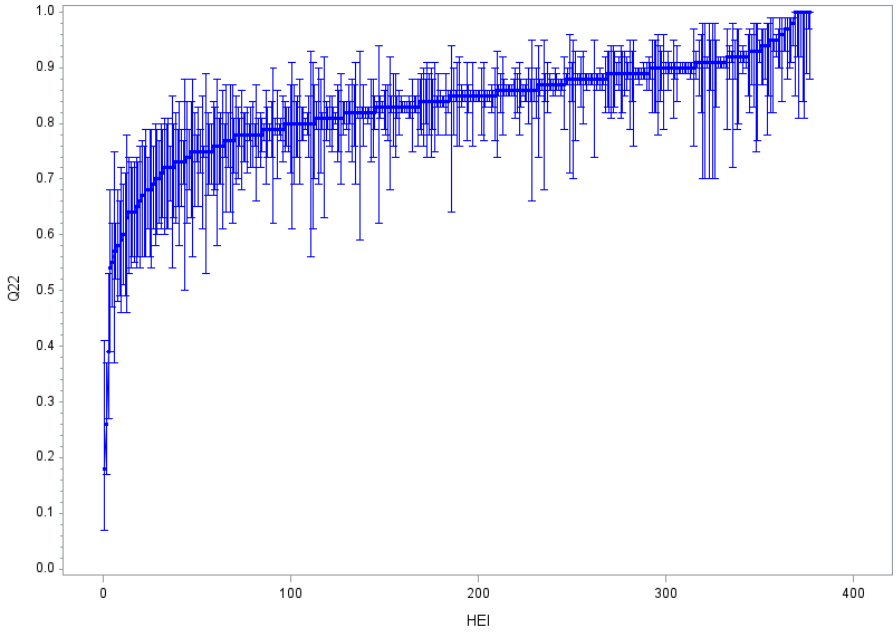

Figure 1 shows that the distribution of responses to question 22 is relatively flat, apart from a few outliers at either end. For the main body of institutions in the middle of the graph, there does not appear to be any significant difference in the outcomes. It may be possible to separate institutions at the extremes of the distribution, those institutions with particularly low and particularly high proportions. However, care would be needed in determining any such thresholds, to ensure the differences are significant. A relatively similar distribution is seen for other questions on the survey.

It should be noted that the limitations of using the raw data for comparisons has been recognised before and the intention is to use comparisons with benchmarks to compare institutions instead[1]. This aspect of the review work has not been completed; it will be covered in the final version of the report. This will aim to build on previous work of the performance indicators steering group to carry out a methodological review of the creation and use of benchmarks.

Confidence intervals are not available for breakdowns of NSS data by sub-groups such as ethnicity and socio-economic classification. Using NSS data at HEFCE, the review was able to examine the sample sizes for different sub-group categories to get an indication of likely precision. Generally speaking, when breaking down response categories into a single classification sample sizes are of a reasonable size (mostly above 50 responses), but given the confidence intervals at the overall level it is inevitable that comparisons of raw data between institutions at this level would not be statistically significant. Whether this would also be an issue for the benchmarking approach would need to be examined.

When combining more than one classification, cell sizes quickly become too small to be of any use for statistical inference. Generally speaking, when looking at individual institutions there is a wider variation in numbers for particular sub-groups, especially for subjects.

Details on a few specific cases of unit response in the 2015 NSS data are given below.

[1] Information on benchmarks can be found at www.hesa.ac.uk/pis

Ethnicity

The smallest cell size for a response category (1 to 5) is 127 at the national level. Regional breakdowns also have reasonable cell sizes, the smallest English region x ethnic category having a response of 55.

Courses

When breaking down results by the main subject level classification, which has 107 categories, there are some very small cells for a few categories. The smallest cell had only 5 responders and there were 15 cells with fewer than 50. However, the small cells were for very minor subjects. For most subjects the cell sizes were fine on their own, but many would be too small is broken down into further cross-classifications.

HEFCE had separately analysed courses in DLHE destinations data and NSS Q22 data. For DLHE, 36% of courses were too small to publish, for NSS 34% of courses were too small to publish. The situation is much worse if concentrating on part-time students - in this case around 80% of courses were too small to publish. For NSS, there are much smaller cell sizes in general for part-time students.

Conclusions

The analysis above shows that comparisons between raw data for institutions at the overall level are not usually significant, with the exception of a small number of outliers at each end of the distribution. Breaking down data by a single categorical variable (for example, by age or sex) generally gives large enough sample sizes for general statistical analysis (but not for comparisons between institutions), but breaking down data by more than one variable leads to insufficiently large sample sizes. Pooling data across multiple years would improve the capacity for statistical analysis.

A short review of the effectiveness of adopting a comparison of institutions against benchmarks will be included in the final version of the report.

4 Conclusions

4.1 Data Collection and Collection Instruments

The HESA collection for the student record and the NSS and the DLHE collection processes are all well-established. Their target populations are clearly defined and the processes for collecting the data are effective. The HESA collection is a very substantial exercise and the systems for checking the data supplied by HEIs are comprehensive. Feedback from two HEIs describes the exercise as challenging and significant effort goes into providing the data. ONS has previously assessed the quality of the student record data and found it to be very good.

A review of the NSS and DLHE collection instruments is not yet complete; it will be covered in the final version of the report.

4.2 Coverage

Although each of the HESA student record, the NSS and the DLHE have clearly defined target populations, these are not necessarily what is required for the TEF. At the time of this review, the TEF scope hasn't been specified to the review team. It is important that this is clearly specified. Without it, it is not possible to assess the degree to which the contributing data sources match the desired population.

Recommendation 1: Define the target population for the TEF

With the target population identified, the data sources can be assessed to identify the degree to which they match the TEF target population. If the comparison shows overcoverage and/or undercoverage, it may be possible to revise the data collections. Where there remain mismatches, a standard approach would see weighting applied to adjust for bias.

Recommendation 2: Determine the extent of under and overcoverage from the data sources. Modify the coverage of the data sources if possible and determine weightings to account for remaining differences.

4.3 Non-response

The response rates for the NSS and the DLHE are good by modern standards for non-compulsory surveys. However, there is still potential for bias where the characteristics of non-responders differ from responders. A limited analysis of non-response has identified differential response for a number of key population characteristics.

Recommendation 3: Further analysis of the characteristics of responders and non-responders should be carried out. If differences are found, determine weights to adjust for the differences. One way of exploring this is to follow-up on non-responders and to persuade them to fill-in the questionnaires; this is usually not easy. However, with HESA student record data, there is already a rich source of student data available to use to identify whether characteristics differ between responders and non-responders.

Traditionally, weights can be used to adjust for differences between responders and non-responders.

4.4 Potential for inference

For the NSS results, this review has briefly considered the degree to which valid comparisons can be made between HEIs. The publication of confidence intervals on the student responses for each question is helpful.

When using raw data, in most cases, the differences between institutions at the overall level are small and are not significant. It may be possible to identify a small number of institutions which are significantly different, that is significantly better or worse. Breaking down data by a single categorical variable mostly gives sufficient sample sizes for some statistical analysis (but not comparisons between institutions). Breaking down data by more than one variable leads to insufficient sample sizes.

The TEF metrics will be comparing institutions' performance to benchmarks that account for some of the characteristics of their student intake. Analysis will aim to show whether institutions are significantly above or below their benchmark and there comparisons will be used in TEF assessments. TEF judgements will be based on the benchmarked metrics and qualitative provider submissions. The methodology for creating the benchmarks, comparing to individual institutions' performance assessing statistical significance is an important element and is part of a fundamental review of the UKPI's benchmarking approach [15] by the UK Performance Indicators Steering Group.

A short review of benchmarks will be included in the final version of the report.

4.5 Stakeholder Engagement

The limited engagement with stakeholders involved in providing and using the data has shown a degree of doubt over the use of the data for assessment purposes. Continuing engagement with data providers will be needed.

It is recognised that BIS already consulted stakeholders by launching the Higher Education Green Paper consultation – "Higher education: teaching excellence, social mobility and student choice" in November 2015 and will be launching a technical consultation on the metrics shortly.

Recommendation 4: Continue to engage with data providers and users to ensure their views and concerns are captured and addressed.

4.6 Overall Comments

This review has assessed three current data sources against general quality standards used in official statistics. A number of aspects the data collection and statistical processing have been identified where improvements are recommended; these would improve the quality of the data for the purposes of contributing to the TEF.

The following table summarises the recommendations and provides a priority rating:

Number	Recommendation	Priority
1	Define target population	High
2	Determine extent of under and over-coverage	High
3	Non-response analysis and treatment	High
4	Stakeholder engagement	Medium

References

1. The Teaching Excellence Framework: Assessing quality in Higher Education, http://www.publications.parliament.uk/pa/cm201516/cmselect/cmbis/572/572.pdf
2. Guidelines for Measuring Statistical Quality, ONS, www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/index.html
3. The Generic Statistical Business Process Model, UNECE, http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model
4. Sanderson R. and Bremner C. (2015), Improving quality through regular reviews: implementing regular quality reviews at the Office for National Statistics, Survey Methodology Bulletin No 74, www.ons.gov.uk/ons/guide-methodology/method-quality/survey-methodology-bulletin/index.html
5. Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*, 77-100.
6. Willis, G.B, (1999) 'Cognitive Interviewing: A 'how to' guide', Short course presented at the 1999 Meeting of the American Statistical Association.
7. Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research, science research. The Qualitative Report, 12, No 2, 281-316.
8. In House Standards and Guidance for paper questionnaires, ONS, internal document written by the ONS Data Collection Methodology team.
9. Dillman, D et al (2014), Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method
10. Review of Quality Assessment, HEFCE 2016, www.hefce.ac.uk/lt/nss/Research/
11. Review of the National Student Survey, NatCen Social Research, July 2014, http://www.natcen.ac.uk/our-research/research/review-of-the-national-student-survey/
12. Beyond 2011: Administrative Data Sources Report: HESA Student Record, May 2013, www.ons.gov.uk/census/censustransformation programme/beyond2011censustransformationprogramme/
13. Heaney, N. "Analysis of students excluded from the National Student Survey", HEFCE, October 2015
14. Ritchie, J., Spencer, L., and O'Connor, W. (2003) Carrying out qualitative analysis. In Ritchie, J. and Lewis, J. (eds.) *Qualitative research practice: A guide for social science students and researchers*. Sage Ltd, London.
15. See paragraph 6.2 https://www.hesa.ac.uk/dox/performanceIndicators/PISG/UKPITG_2015_07/UKPITG_minutes_Jul_2015.pdf

Appendix - Methodology and processes behind the cognitive interviewing process
Interview technique and procedure

Cognitive interviews were conducted in order to ensure that interviewers collated as much information as possible. Interviewers also observed respondents as they completed the questionnaire making note of any behaviours they deemed relevant e.g. satisficing or signs of frustration while completing the on-line version of the questionnaire. Cognitive interviewing is a recognised method of testing questions and questionnaires, which uses techniques such as paraphrasing, think aloud, and concurrent and retrospective probing. It aims to provide an explanation of concepts, processes and patterns of interpretation [5]. It was decided that face to face interviews were the most suitable technique for gathering data given the complexity of the information that was being gathered.

Analysis

All interviews were audio recorded to enable a more thorough analysis. The interviews were transcribed verbatim. The data from the interviews were organised and sorted into themes to identify patterns across the dataset. This method facilitates in the systematic identification and examination of related data [11].

For expediency all interviews were transcribed directly into an analysis framework the design of which related directly to the questions and probes that were used during the interview process. Salient behaviours noted by the interviewers were also recorded in the relevant section of the analysis chart. Thematic analysis was then used to extract any themes that arose from the data obtained. In support of the themes extracted from the data obtained, any relevant quotes were added 'verbatim' to the relevant section of the framework.

It should be noted that qualitative analysis does not allow the researcher to draw quantitative conclusions so statistical reliability cannot be inferred from the results obtained. Additionally, while the respondents are representative of the sampling criteria one cannot discount the possibility that some responses are unique and therefore cannot be extrapolated to the wider population. That said, where opinions recur through the data it is possible to suggest that the responses obtained are more likely to be representative of the wider population than not.

Confidentiality and ethical issues

Respondents' details and all information that was provided were treated in confidence according to ONS guidelines and DCM protocol, which are based on UK Government Social Research (GSR) and Social Research Association (SRA) guidelines. All participants were informed of the rationale behind the interview and were advised of the confidentiality of the information that they provided both on-line and to the interviewer and were also advised of their right to withdraw from the interview should they so wish.

Methodology behind expert reviews

ONS often carry out expert reviews as an evaluation method for the surveys they administer. By comparing collection instruments against standard ONS design principles researchers are able to ensure standardisation within the organisation. The same process can be applied when offering expert advice to external clients. Carrying out such a procedure helps to prevent inconsistencies between, and improve reliability within, collection instruments where more than one is used within a given project. The review process also helps to highlight unrecognised complexities in questions and in survey navigation, as well as ensuring the correct design of questions from question and answer option wording to the order in which they should be displayed.

Each question within a questionnaire is evaluated and a re-design is proposed where necessary to make sure that burden on the respondent to be reduced to a minimum and to ensure that the proposed structure is applicable to all respondents within a given sample. Ideally, data from each of the collection instruments should also be available as this allows the researcher to gauge item non-response and accuracy rates within the answers provided. These sets of analyses allow the researcher to focus on areas where completion may be an issue for the respondent, either because of lack of usability or lack of comprehension.

At the end of the review process the client will ultimately receive a report which highlights all potential issues with the design of the questionnaire along with an indication of how these issues can be addressed.